

ONTOEPFS: UMA ONTOLOGIA PARA EXPERIMENTOS EM MAPEAMENTO DIGITAL DE SOLOS

LUIZ MANOEL SILVA CUNHA¹

DANIEL DE OLIVEIRA²

EDUARDO OGASAWARA³

SANDRA FERNANDES ANDRADES⁴

GERALDO ZIMBRÃO DA SILVA⁵

MARIA DE LOURDES MENDONÇA BREFIN⁶

RESUMO: A descrição formal de experimentos para aquisição, tratamento, análises de dados, compartilhamento e reprodução dos resultados obtidos, é fundamental para os cientistas envolvidos. Ontologias vêm sendo utilizadas para criar modelos formais para disseminação de informação e conhecimento. Este artigo apresenta uma ontologia para apoiar o Experimento de Predição de Informações de Fertilidades de Solos (OntoEPFS) em Mapeamento Digital de Solos (DSM). O EPFS se encontra em desenvolvimento na Empresa Brasileira de Pesquisa Agropecuária, no Centro Nacional de Pesquisa em Solos – Embrapa Solos. A adoção dessa ontologia, por parte da Embrapa Solos, facilitará a descrição, a formalização e a disponibilização de conhecimentos sobre os experimentos realizados por esta unidade.

Excluído: .

PALAVRAS-CHAVE: workflow científico, ontologia, automação de processos, predição de fertilidade dos solos.

ONTOEPFS: AN ONTOLOGY FOR EXPERIMENTS IN DIGITAL SOIL MAPPING

ABSTRACT: The formal description of experiments for the acquisition, processing, data analysis, sharing and reproduction of these results is important for scientists involved in experimental issues. This paper presents ontology to support the prediction information of fertility of soils experiments in Digital Soil Mapping (DSM). The EPFS is in development in the Brazilian Agricultural Research Corporation in the National Center for Research in Soil - Embrapa Solos. The adoption of the OntoEPFS by Embrapa Solos will enable the description, organization and availability of knowledge about the experiments conducted by this unit.

KEY-WORDS: scientific workflows, ontology, processes automation, prediction soils fertility.

Excluído: (Não tenho condições de fazer a revisão no texto em inglês)†

Formatado: Português

¹ M.Sc. Engenharia de Software, Embrapa Informática Agropecuária, E-mail: luizm@cnptia.embrapa.br

² M.Sc. Engenharia de Sistemas e Computação, PESC/COPPE/UFRJ, E-mail: danielc@cos.ufrj.br

³ M.Sc. Engenharia de Sistemas e Computação, PESC/COPPE/UFRJ, E-mail: ogasawara@cos.ufrj.br

⁴ M.Sc. Geoquímica Ambiental, Universidade Federal Fluminense, E-mail: sfernandesdeandrade@yahoo.com.br

⁵ Pós-Doutorado Sistema de Computação/Banco de Dados, PESC/COPPE/UFRJ, E-mail: zimbrão@cos.ufrj.br

⁶ Pós-Doutorado Sistema de Informação Geográfica e Sensoriamento Remoto, Embrapa Solos, E-mail: loumendonca@cnps.embrapa.br

Excluído: .

Excluído: .

Excluído: .

1. INTRODUÇÃO

Mapeamento Digital de Solos trata-se de um processo utilizando técnicas da matemática, da estatística e de outras áreas do conhecimento para a produção de dados georreferenciados do solo (mapas digitais de solo), com o apoio de recursos computacionais, utilizando vários tipos de dados (solo, meio ambiente, tempo, localização geográfica, etc) (McBratney et al. 2003). Em DSM, vários sub-processos são executados nas formas encadeada e não encadeada, combinando os vários tipos de dados e técnicas das áreas de conhecimento citadas. Eles são suportados por várias tecnologias de informação e recursos computacionais (banco de dados, rede de computador, Internet, Sistema de Informação Geográfica, etc) e, executam ações relacionadas à captura, ao tratamento, à análise, à disponibilização e ao compartilhamento dos dados obtidos. Uma vez que esses sub-processos apresentam essas características, podemos intuitivamente visualizá-los como experimentos científicos *in silico* como proposto por Oinn et al (2007). Esses experimentos lidam com muitas combinações de programas e fontes de dados e são executados completamente em ambientes computacionais de modo automatizado. Cada uma dessas combinações, descritas pelo um fluxo de execução explícito, é chamada de workflow científico (Taylor et al. 2006). Os workflows científicos também provêem facilidades, ao responsável pelo experimento, na construção, na reutilização, análise de dados e no compartilhamento de seus próprios resultados e de seus workflows elaborados. Os workflows científicos são executados por Sistemas Gerenciadores de Workflows Científicos (SGWf). Existem diversos SGWf disponíveis que provêem algumas dessas facilidades (Altintas et al., 2006, Callahan et al. 2006, Oinn et al., 2007, Taylor et al., 2007). No entanto, eles tratam mais da execução individualizada dos workflows do que diz respeito à eficiência, programação e gestão de dados. A falta de apoio em semântica para auxiliar o usuário na concepção, definição e uso de workflows é uma deficiência apresentada por esse tipo de sistema. Os autores acreditam que ontologias oferecem uma solução para esse problema, fornecendo um arcabouço conceitual comum para representação de informação e conhecimento. Ontologias podem ser acopladas a SGWf, fornecendo semântica e apoio na composição, execução e análise de experimentos *in silico*. Diferentes domínios do conhecimento consideram diferentes definições para ontologias (Gomez-Perez et al., 2004). Na ciência da computação, o principal foco está para o qual a ontologia foi projetada. Em muitos casos as ontologias são modeladas a fim de proporcionar o compartilhamento e reuso do conhecimento. Espera-se que, pelo uso de ontologias, seja possível desenvolver um modelo de referência para descrição de experimentos em DSM. Este trabalho apresenta uma proposta de ontologia para descrição de experimentos de predição informações de fertilidade dos solos (EPFS) em DSM. Ela está inserida dentro de um objetivo maior que é disponibilizar, na Web, um sistema capaz de facilitar a gestão de experimento científico para o domínio de DSM, uma lacuna ainda pouco explorada pelos núcleos de desenvolvimento de sistemas de apoio à pesquisa da Embrapa. Na seção 2, é dada uma visão geral do EPFS, suas etapas e como elas se relacionam. Em seguida, seção 3, conceitos e benefícios de uso das ontologias são apresentados. A descrição desenvolvimento da OntoEPFS e a sua versão inicial seguido das conclusões finalizam este trabalho.

Formatado: Fonte: Não Negrito, Cor da fonte: Automática

Excluído: a

Excluído: t

Formatado: Cor da fonte: Automática

Excluído: t

Excluído: através de

Formatado: Fonte: Não Negrito

Excluído: prov

Excluído: ê

Excluído: eem

Excluído: através dos

Formatado: Fonte: Não Negrito

Excluído: prov

Excluído: ê

Excluído: eem

Formatado: Fonte: Não Negrito

Formatado: Fonte: Não Negrito

Excluído: t

Excluído: através do

Formatado: Fonte: Não Negrito

Excluído: s

Excluído: s

2. EXPERIMENTO DE PREDIÇÃO DE INFORMAÇÕES DE FERTILIDADE DOS SOLOS

A Figura 1 apresenta a visão geral do workflow científico do EPFS, contendo as principais atividades. Este workflow foi formalizado após discussão com técnicos responsáveis pela

execução do experimento. O *workflow* científico é composto das seguintes atividades: (i) na preparação de dados, que inclui as tarefas de extração dos registros dos perfis dos solos, contidos no banco de dados da Embrapa Solos, de seleção dos perfis e de interpolação dos dados que descrevem os horizontes, por camadas; (ii) na predição de variáveis regionais, que pode ser decomposta pela aplicação de estatísticas para análise descritiva dos dados dos perfis, pela etapa de eliminação daqueles que estão fora do padrão definido e pela tarefa de predição das variáveis de fertilidades para uma dada região. (iii) no estabelecimento do modelo de predição de fertilidade dos solos a ser utilizado pela atividade de mapeamento digital de fertilidade dos solos; (iv) na recomendação de adubação dos solos (ainda não automatizada).



Figura 1 – Workflow Científico do EPFS.

Durante a análise do EPFS, verificou-se a necessidade de se estabelecer um formalismo para descrição dos experimentos executados em DSM. Foi observada a utilização de vários tipos de *softwares*. Informações fundamentais, tais como, a(s) metodologia(s) empregada(s), os dados utilizados, os recursos computacionais empregados e os resultados obtidos nem sempre são registrados com clareza. A forma e a descrição do armazenamento dos resultados obtidos dos experimentos nem sempre permitem a recuperação e a disponibilização dos mesmos. Isso dificulta, também, a reprodução dos experimentos. Reprodutibilidade é um requisito fundamental de experimentos *in silico*, pois atesta que o processo seguiu um método científico. Tomando os fatos citados como problemas observados, propomos a construção de uma ontologia como um modelo conceitual de EPFS. Acredita-se que com a modelagem e, a aplicação efetiva desta ontologia, alguns dos problemas citados possam ser sanados ou amenizados. Na próxima seção são descritos conceitos, tipos, benefícios e classes de ontologias.

3. ONTOLOGIAS: CONCEITOS E APLICAÇÕES

Para Gruber (1993), uma ontologia é “uma conceituação compartilhada de um determinado domínio”. Gomes-Perez (1999) define as ontologias como “conjuntos de termos ordenados hierarquicamente para descrever um domínio que pode ser utilizado como um arcabouço para uma base de conhecimento”. Outros conceitos para ontologias são encontrados em Guarino e Welty (2000) e, o seu significado preciso tende a variar conforme o objetivo do uso da ontologia. Como benefícios, elas apresentam: (i) a formalização da terminologia do domínio; (ii) facilidades para compartilhamento de conhecimentos do domínio; e (iii) fornecimento de uma descrição exata do conhecimento. Segundo Gruber (1993), as ontologias são formadas por (i) classes – usadas com ampla abrangência, podendo modelar qualquer coisa sobre a qual alguma coisa é dita, como uma tarefa, função, ação, entre outras; (ii) relações – representam os tipos de interação entre as classes e o domínio, por exemplo, “subclasse de” ou “conectado

a” e; (iii) axiomas – usados para modelagem de sentenças que são sempre verdadeiras. De acordo com Guarino e Welty (2000), as ontologias podem ser classificadas em: (i) ontologias de alto-nível; (ii) ontologias de domínio; (iii) ontologias de tarefa; e (iv) ontologias de aplicação. As ontologias podem ser compartilhadas para uso, em conjunto com outras ontologias ou ferramentas que possibilitam também a interoperabilidade.

4. A ONTOEPFS: UMA ONTOLOGIA PARA EXPERIMENTOS DE PREDIÇÃO DE FERTILIDADE DE SOLOS

Os trabalhos descritos em (Oliveira et al. 2008, 2009) serviram de base para formulação da OntoEPFS. Eles abordam questões sobre definição de metadados, metodologia e relatam experiências envolvendo construção de ontologias. Também foram utilizados os conhecimentos e as experiências de especialistas na área de DSM.

Com base no diagrama da Figura 1, foram estabelecidas as classes e suas respectivas subclasses consideradas principais para a ontologia, são elas: (i) reg_conclusões_exp - faz referências às conclusões obtidas com o fim dos experimentos; (ii) dado_exp - faz referência aos dados utilizados pelo experimento; (iii) desc_info_geral_exp - faz referência às informações que caracterizam os experimentos; (iv) resultados_exp - faz referência aos resultados obtidos ao final da execução dos experimentos; (v) passos_exp - faz referência aos passos seguidos pelos experimentos; (vi) algoritmos_exp - faz referência aos algoritmos empregados nos experimentos; (vii) metodos_exp - faz referência a cada um dos métodos implementados pelos softwares; (viii) softwares_exp - faz referência aos softwares utilizados na execução dos experimentos. A Figura 2 exibe as classes da OntoEPFS através do software Protege (<http://protege.stanford.edu/>).

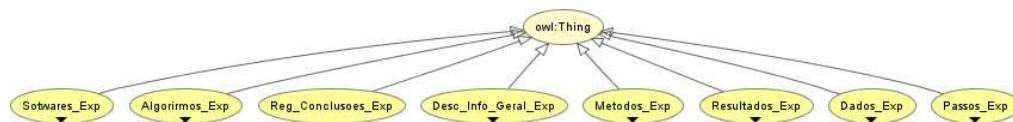


Figura 2 – Parte das classes e subclasses da OntoEPFS.

5. CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho foi apresentada a OntoEPFS como uma ontologia para descrição formal dos experimentos científicos em DSM. Essa ontologia vem sendo desenvolvida juntamente com EPFS, na Embrapa Solos. Até a conclusão desse artigo, consultas ao Acervo Documental da Embrapa (www.bdpa.cnptia.embrapa.br), hoje, composto de 492.764 registros sobre trabalhos técnicos desenvolvidos, nenhum documento foi encontrado relatando experiências relacionadas ao trabalho proposto. Dessa forma, pode-se dizer que esta é uma iniciativa inovadora na Embrapa.

Uma vez concluída e implantada a ontologia, espera-se disponibilizar um instrumento capaz de (i) auxiliar na estruturação de uma base de conhecimentos sobre experimentos científicos em DSM envolvendo as etapas apresentadas na Figura 1; (ii) evitar que experimentos com características semelhantes sejam totalmente desenvolvidos, fomentando a reutilização de experimentos anteriores. Com isso não ocorre o retrabalho e vários gastos desnecessários (humano, tempo, financeiro, etc); (iii) facilitar a recuperação de informações sobre os experimentos para fins de emissão de relatórios sobre os mesmos.

Para alcance desses benefícios, estão previstos como trabalhos futuros: (i) validação da OntoEPFS junto aos cientistas da Embrapa Solos; (ii) aplicação do método OntoClean (Guarino e Welty, 2002) e (iii) verificação se a OntoEPFS é capaz de responder as seguintes

Excluído: t

Excluído: ,

Excluído: ,

Excluído: re-

perguntas (i) dado um serviço de mapeamento digital de solo, que outros serviços são equivalentes a este, ou seja, implementam a mesma tarefa? (ii) dada uma tarefa no *workflow* EPFS, quais são os dados de entrada e saída? (iii) dadas às atividades α e β , a serem executadas seqüencialmente no *workflow* EPFS, os tipos de dados de saída de α são compatíveis com os tipos de dados requeridos por β ? Essa lista poderá ser aumentada à medida que a OntoEPFS seja aperfeiçoada.

Excluído: seq

Excluído: ü

Excluído: uencialmente

Excluído: t

6. REFERÊNCIAS

ALTINTAS, I.; LUDAESCHER, B.; KLASKY, S.; VOUK, M. A. Introduction to scientific workflow management and the Kepler system. In: CONFERENCE ON HIGH PERFORMANCE NETWORKING AND COMPUTING, 2006, Tampa, Florida. *Proceedings of the 2006 ACM/IEEE conference on Supercomputing*. New York: IEEE/ACM, 2006. (Article n°. 205).

CALLAHAN, S. P.; FREIRE, J.; SANTOS, E.; SCHEIDEGGER, C. E.; SILVA, C. T.; VO, H. T. VisTrails: visualization meets data management. In: INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA, 2006, Chicago, Illinois. *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*. New York: ACM, 2006. p. 745-747.

GOMEZ-PEREZ, A.; CORCHO, O.; FERNANDEZ-LOPEZ, M. *Ontological engineering: with examples from the areas of knowledge management, e-commerce and the semantic web*. London: Springer, 2004. 415 p. (Advanced information and knowledge processing).

GRUBER, T. R. A translation approach to portable ontology specifications. *Knowledge Acquisition*, v. 5, n. 2, p. 199-220, 1993.

GUARINO, N.; WELTY, C. Evaluating ontological decisions with OntoClean. *Communications of the ACM*, v. 45, n. 2, p. 61-65, Feb. 2002.

GUARINO, N.; WELTY, C. A. A formal ontology of properties. In: DIENG, R.; CORBY, O. (Ed.). *EKAU '00: proceedings of the 12th European Workshop on Knowledge Acquisition, Modeling and Management*. London: Springer-Verlag, 2000. p. 97-112. (Lecture Notes in Computer Science, v. 1937).

MCBRATNEY, A.; SANTOS, M.; MINASNY, B. On digital soil mapping. *Geoderma*, v. 117, p. 3-532, 2003.

NOY, N.; MCGUINNESS, D. *Ontology development 101: a guide to creating your first ontology*. Disponível em: <ftp://ftp.ksl.stanford.edu/pub/KSL_Reports/KSL-01-05.pdf.gz>. Acesso em: 12 maio 2009.

OINN, T.; LI, P.; KELL, D. B.; GOBLE, C.; GODERIS, A.; GREENWOOD, M.; HULL, D.; STEVENS, R.; TURI, D.; ZHAO, J. Taverna[™]Grid: aligning a workflow system with the life sciences community. In: TAYLOR, I. J.; DEELMAN, E.; GANNON, D. B.; SHIELDS, M. (Ed.). *Workflows for e-Science: scientific workflows for grids*. London: Springer, 2007. p. 300-319.

OLIVEIRA, D.; BAIÃO, F.; MATTOSO, M. *MF-Ontology: an ontology for the text mining domain*. Rio de Janeiro: COPPEUFRJ, 2009. 26 p. (Technical Report - ES-727/09). Disponível em: <http://www.cos.ufrj.br/uploadfiles/1234273167.pdf>. Acesso em: 12 maio 2009.

Formatado: Português

OLIVEIRA, D.; BAIÃO, F.; MATTOSO, M. MF-Ontology: uma ontologia para o processo de mineração de textos. In: SEMINÁRIO DE PESQUISA EM ONTOLOGIA NO BRASIL, 1., 2008, Niterói. *Trabalhos aprovados – final*. Disponível em: <http://www.uff.br/ontologia/artigos/311.pdf>. Acesso em: 12 maio 2009.

PINTO, H. S.; MARTINS, J. P. Ontologies: how can they be built? *Knowledge and Information Systems*, v. 6, n. 4, p. 441-464, July, 2004.

TAYLOR, I. J.; DEELMAN, E.; GANNON, D. B.; SHIELDS, M. (Ed.). *Workflows for e-Science: scientific workflows for grids*. London: Springer, 2007. 523 p.

TAYLOR, I.; SHIELDS, M.; WANG, I.; HARRISON, A. The Triana workflow environment: architecture and applications. In: TAYLOR, I. J.; DEELMAN, E.; GANNON, D. B.; SHIELDS, M. (Ed.). *Workflows for e-Science: scientific workflows for grids*. London: Springer, 2007. p. 320-339.

Formatado: Inglês (EUA)